DOCUMENT RESUME

ED 122 605 FL 007 566

AUTHOR Boven, J. Donald
TITLE An Experimental Integrative Test of English Grammar. Workpapers in Teaching English as a Second Language, Vol. 9.
INSTITUTION California Univ., Los Angeles. Dept. of English.
PUB DATE Jun 75
NOTE 16p.

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *English (Second Language); Grammar; Language Instruction; Language Proficiency; Language Skills; *Language Tests; *Linguistic Competence; *Listening Comprehension; Listening Tests; Second Language Learning; *Standard Spoken Usage; Testing

ABSTRACT

The experimental test measures the ability of a subject to reconstruct obscured words by means of sentence analysis, carried out not as a separate academic task, but in the normal procedure of understanding what the sentence says. It is a task built on the assumption that the ability to handle reduced redundancy is a valid measure of linguistic competence. The reduced redundancy in this type of test is a consequence not of deliberate deletions or masking by superimposed noise, but of the reductions, assimilations, and contractions that normally accompany sentence production by native speakers functioning in a relaxed, informal context. The test is easy to administer and to score and is highly objective. It clearly separates nonnative from native speakers and spreads the nonnatives over an unusually wide range of scores. Performance on the test appears to correlate well with other measures of competence. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED122605

AN EXPERIMENTAL INTEGRATIVE TEST OF ENGLISH GRAMMAR

J. Donald Bowen

An interesting and provocative discussion has recently been carried on by two of my colleagues, and others interested in testing and evaluation, about the relative merits of discrete-point and integrative tests of language competence.[1] I don't wish to enter this arena as a proponent of one or the other side of the debate, since I believe the problems of language testing are so extensive that we should encourage the contributions of as many scholars as can be interested in the subject of testing, and should further encourage them to follow any philosophy or viewpoints they feel might yield useful results. When the issues are clarified, then maybe we can build a better test with the information that the discussions will have generated.

The issues are well-known and will be repeated here only in summary form to provide a context for further comment. The discrete-point test can be focussed on the specific learning problems of a student, reflecting his background and target languages. The information such a test provides is immediately useful for diagnostic and remedial purposes. The test is easy to administer, easy to score, and reliable. But it may require special skills for its construction and is often difficult to interpret, since appropriate weightings for different items and sections must be established before a general assessment of competence can be made.

Integrative tests, on the other hand, test indeterminate bundles of skills such as are presumed to be involved in normal communication. Such tests are easier to construct and easier to interpret, and being more like the activities of lesson presentation do not interrupt in so conspicuous a way the normal classroom routine. John Oller claims that "scores on tests of integrative skills tend to correlate better with teacher judgments, better among themselves, and better with other measures of language skills than do any of the discrete-point types because they more nearly reflect what people actually do when they use language."[2] Oller may be right, or Rand may be right that we are fooling ourselves with the manipulation of data, or more likely both of them are partly right and the best answer lies somewhere between the polar positions they seem to represent.[3] I maintain there is a need for further study and that research in the area of language-competence testing is especially welcome. We are still a long way from the kind of results that would justify full confidence in our pronouncements. Indeed, if our testing techniques were sufficiently well established, it might be possible to spend less time quibbling about the relative merits of this or that application of methodological theory. Verified, measured results would be quickly established by valid, reliable data, and these would serve as measures of the relative value of various contrasting applications.

FL007566

In the meantime it would seem that certain kinds of tests favor certain applications--some more appropriately assessed by integrative tests, some by discrete-point. Indeed the definition of difference suggests that while to some extent either approach is possible in most skill areas, integrative tests are readily adaptable as measures of aural and written comprehension, reading,

composition, and discrete-point tests are well adapted to assessing the pronunciation of specific sounds, individual vocabulary items, and specific points of grammar, such as concordance patterns.

Integrative tests indeed are defined as measures in which if something goes wrong it is difficult or even impossible to point out precisely what points of language use are involved. To show this contrast, let's examine two items, one discrete-point and one integrative. The instruction for the first item could be "Choose the correct form for the blank."

1. My little brother _____ at home right now.

   _____ A. was

   _____ B. is

   _____ C. being

   _____ D. are

   _____ E. been

The answer has to be is, none of the other options will fit. The instructions for the second item will have to be "Choose the best paraphrase for the following sentence."

2. Sitting on the floor, my sister Karen watched her husband paint the chair.

   _____ A. While Karen sat on the floor, her husband watched her paint the chair.

   _____ B. While she sat on the floor, my sister watched a chair being painted.

   _____ C. While her husband watched her sitting on the floor, Karen painted the chair.

   _____ D. Karen watched her husband paint the chair that was sitting on the floor.

   _____ D. My sister watched the floor while her husband was painting the chair she was sitting on.

The first item must take option C because of a rule of English grammar which says verbs will agree with their subjects in person and number. The second item takes option B because it expresses a message consistent with that of the stem. No single grammar rule can be cited in justification. Note that option D might be considered a possible correct answer, but it's not the best one because it repeats fewer details than option B and makes an assumption that was not stated, namely that the chair is "sitting on the floor." It might well be on a workbench. The second item does not really concern grammar; it tests comprehension or usage.

Is it possible to construct a grammar test that is primarily integrative?

An attempt to do so is reported in the remainder of this paper. The test is built on an assumption expressed very clearly by Richard R. Lee when he points out that "the ability to reconstruct a distorted speech signal distinguishes the native speaker from even the advanced non-native speaker."[4] I further assume that this type of procedure can identify and rank non-native speakers of different levels of competence.

In various experiments in the past the "distortion" has taken the form of omissions from the written text (employing the familiar cloze test format), or of similar omissions from the spoken text (physically replacing recorded segments of tape with blank tape), or by an overlay of white noise that interferes with a recorded message (so that the task of understanding resembles listening to a telephoned message when there is excessive line noise). These externally imposed distortions force the listener to rely on redundancy patterns in order to interpret the message. The more accurately the person tested can cope with a disfigured or partially covered message, the better his overall competence in the language.

In the test presently proposed and described the "distortion" is not imposed from factors beyond the language itself, but from within. Indeed it may be inappropriate to refer to normal modifications as distortions, but they are occasionally so considered by teachers who limit the classroom use of English (the language used to illustrate the test) to specifically formal levels of speech. I refer to the systematic modification of sounds and sound sequences in patterns of contraction and assimilation, where the citation forms of certain words are "obscured" or modified in ways that affect the identification (and for the uninitiated the intelligibility) of individual items in the stream of speech. Such obscured forms must then be interpreted by reference to other items or parts of the sentence (i.e., through cooccurrence patterns, contextual constraints, anaphoric reference, etc.)

In passing one might note that such internal "distortion" may be pedagogically more useful than cloze tests, mutilated recordings, noisy channels, etc. because the test material comes from normal use of the language. Mastery of these modifications would certainly enhance the ability of a second-language student to interact with native speakers in normal communication situations. Indeed the idea for the test format first occurred to me as a drill, and it is in my opinion quite possible that drill is the most useful and productive application of the idea--a means to improved listening comprehension.

In the test the task is to reconstruct certain obscured words by means of sentence analysis. As examples consider the following:

3. Here 't seems we're always busy.

4. Here 't headquarters we're always busy.

The first two words "here 't" are interpreted "here it" in item 3 but "here at" in item 4, the determining information being the form of the word which follows: the verb "seems" requires the interpretation it and the noun "headquarters" requires at. There is no cue in the actual pronunciation, which in both cases is /hírət/.

The task set for the student is to identify and write down the full form of the second word in each sentence. The test, then, presumes a knowledge of the written forms of English. This is not a serious limitation, since the level of competence necessary to take the test is sufficiently high that it would be a rare subject who could make the identifications but not know the necessary amount of English spelling to write down the words specified. Here are three additional examples of cooccurrence requirements:

5. Who'd he been to see?

6. Who'd he want to see?

7. Who'd he like to see?

The verb forms "been, want, like" require that the contraction 'd in /húwdiy/ be interpreted "had, did, would" respectively. A fully competent speaker of English knows this immediately; a second-language student depends on the degree of his familiarity with these verbal forms, which is no doubt part of what is being tested.

One final example:

8. Give 'm an inch and he'll take a mile.

9. Give 'm an inch and they'll take a mile.

The contracted form 'm is "him" in 8 and "them" in 9, information which is conveyed by the appearance of the pronouns "he" and "they" subsequently in each sentence.

The test, to be copied below, consists of 50 items, recorded on tape for standardized administration. It was first thought that the order and relationship of items might have an effect on performance, so suggestively similar items were placed in sequence on one version of the test. For example, contractions involving has and is (actualized as /z/) were interspersed, but placed together, followed by a set involving to, of, have, a (actualized as /ə/) again interspersed but placed in an adjacent sequence, etc. In a second version the total sequence was randomized, with no possible hints suggested by the order of the items. The full test, then, was 100 items, 50 items repeated in different sequences.

The first administration of the test was on October 23, 1974, to two classes of eleventh-grade male students, twenty-four in all, at the Collège de la Sainte Famille in Cairo, one class from each of two streams, one using French and one using Arabic as principal media of instruction.[5] The results, comparing streams, format, and order, were:

	Stream		Version		Order	
	French	Arabic	Sequenced	Random	First fifty	Second fifty
Mean score	35.5	29.6	35.0	34.1	31.5	37.6

These results confirm the assumption that the students in the French stream are academically superior to those in the Arabic stream. This was expected, since instruction in French carries considerable prestige and since admission to the French stream is based on demonstrated linguistic competence in French, which correlates with general academic achievement. But there was no significant difference in performance on the sequenced and randomized versions of the fifty test items. Apparently juxtaposition of similar or contrasting items had no effect on performance.

But, perhaps since this was an unfamiliar test activity, there appears to be a practice effect; subjects do better on the second part of the test, regardless of which version is given first. Consequently, in subsequent administrations of the test the sequenced version has been given first and the randomized version second. Scores have been kept separately for the two versions to see if practice (on the sequenced version) indeed had any effect on performance (on the randomized version).

As a result of the pilot test given at the College de la Sainte Famille, a few items were revised, a more elaborate set of instructions was prepared, and the test was re-recorded. The revision of the instructions was prompted by the experience of having a few good students miss items because they added words to the beginning of sentence fragments to make them grammatically complete, and this naturally affected the identification of the second word. The full revised test, including instructions, is given below in the format:

Instructions and Sample Items
Practice Test
Evaluation Test

The complete tapescript follows:

Tapescript for Integrative Grammar Test

In the exercise that follows you are asked to identify and write down on your paper the full form of the second word of each sentence you hear. In some sentences the second word will not be pronounced clearly and separately. It may even be contracted with another word. But you should write down its full form as a single word. So if you hear a sentence like "Who's that?" you should write down the word <u>is</u>, even though all you heard was the /z/ on the end of "who's".

Let's try three practice sentences before we start. Ready? The first will be

Sentence A: Whadja do yesterday?

Write down the full form of the second word.

(five-second pause)

Have you written the word <u>did</u>? You should have, since the sentence, slowed down, would be "What <u>did</u> you do yesterday?" and the second word is <u>did</u>. Try another practice sentence,

Sentence B: Dja finish your work by four o'clock?

Write down the full form of the second word.

(five-second pause)

What word did you write down? It should have been you, since the sentence, slowed down, is "Did you finish your work by four o'clock?" That's not too hard. All you have to do is think of the sentence in a slowed down version and then identify the second word.

One caution. Don't add words to the beginning of the sentence. Just interpret the words that are there. Some sentences are really sentence fragments of types commonly found in spoken English. Here's an example in

Sentence C: Think they'll be back by Friday?

What is the second word?

(five-second pause)

Did you write they? You should have, because they was the second word used, even though the sentence could be expanded to a more complete form "Do you think they will be back by Friday?" But the correct answer is they from "Think they. . ." and not you, from "Do you think they . . ."

OK, ready? Let's go. There will be fifty sentences in the exercise, then another fifty. If you are not sure, guess. But be sure to finish one sentence before the next one comes along, because none of them will be repeated. Or, skip a difficult sentence and listen to the next one. Here we go.

1. Jess 's gone over the accounts already.
2. Just 's not right for him to work all night.
3. He's finished with the first test, isn't he?
4. This 's got to be the best we've ever done.
5. Where's there a pharmacy in this part of town?
6. There's got to be a better way to do this.
7. Just 's I thought; he's not coming.
8. What's been done to improve the class?
9. Whose dancing '11 be judged best?
10. Gonna get here for the soccer match?
11. Cup o' coffee when you have a minute?
12. Coulda been the postman ringing the doorbell.
13. Wanna rain like the one we had last week?
14. Wanna go with me to the movies tonight?
15. What a nice drink of orange juice that was.
16. What 'n old fool he is to say such a thing.
17. Justice 'n peace: that's what we need.
18. This 'n '11 do the job all right.
19. What 'n the name of heaven are you doing?
20. Want 'n interesting book to read?
21. John 'n Nancy 're coming to the party tonight.
22. Give 'em a few days and they'll be back.

23. Give 'im five minutes and if he doesn't come, leave 'im.
24. Ahmana see 'im in just two weeks.
25. Where 'm I gonna get five thousand dollars?
26. How 'll he ever get finished in time to go at noon?
27. When sh' we tell him to finish the schedule?
28. Izzyer brother coming for the graduation?
29. Wouldja like to have the party at my house?
30. Let cher brother take your place if you can't come.
31. So there 're no more books in the storeroom.
32. So they're gonna come after all.
33. Is their car the one that was in an accident?
34. What they're saying shouldn't be repeated.
35. Bill 'r Ralph 'll come, but the others can't.
36. Whose 're those keys on the table?
37. Saw 'er come in just a minute ago.
38. D' he say what time we're supposed to meet tonight?
39. Willy said just what I told him to.
40. Will 'e say what I told him to?
41. Here 't seems like we're always busy doing something.
42. Here 't headquarters it seems like we're always busy.
43. How 'd he ever be able to get here in time?
44. Who 'd he been to see yesterday?
45. Who 'da thought he'd ever remember her?
46. What 'd he do all day at the library?
47. What 'd he done that made the judge so angry?
48. Where 'd you like to go on your vacation?
49. Whattaya think 'll happen to the new president?
50. 'Zat a copy of the original document?

1. Give 'em a few days and they'll be back.
2. Where 'm I gonna get five thousand dollars?
3. What 'n old fool he is to say such a thing.
4. What's been done to improve the class?
5. How'll he ever get finished in time to go at noon?
6. Willy said just what I told him to.
7. Who 'da thought he'd ever remember her?
8. Saw 'er come in just a minute ago.
9. Here 't seems like we're always busy doing something.
10. Wanna rain like the one we had last week?
11. He's finished with the first test, isn't he?
12. 'Zat a copy of the original document?
13. John 'n Nancy 're coming to the party tonight.
14. So there 're no more books in the storeroom.
15. Whattaya think'll happen to the new president?
16. How 'd he ever be able to get here in time?
17. Just 's not right for him to work all night.
18. Justice 'n peace: that's what we need.
19. Wouldja like to have the party at my house?
20. Whose 're those keys on the table?
21. Gonna get here for the soccer match?
22. What 'd he done that made the judge so angry?
23. What 'n the name of heaven are you doing?
24. Jess 's gone over the accounts already.

25. D' he say what time we're supposed to meet tonight?
26. Ahmana see 'im in just two weeks.
27. What 'd he do all day at the library?
28. Bill 'r Ralph 'll come, but the others can't.
29. There's got to be a better way to do this.
30. Here 't headquarters it seems like we're always busy.
31. Cup o' coffee when you have a minute?
32. When sh' we tell him to finish the schedule?
33. Want 'n interesting book to read?
34. Where 'd you like to go on your vacation?
35. This 's got to be the best we've ever done.
36. Is their car the one that was in an accident?
37. Will 'e say what I told him to?
38. Let cher brother take your place if you can't come.
39. This 'n 'll do the job all right.
40. Who'd he been to see yesterday?
41. So they're gonna come after all.
42. Wanna go with me to the movies tonight?
43. Where's there a pharmacy in this part of town?
44. Izzyer brother coming for the graduation?
45. What a nice drink of orange juice that was.
46. Whose dancing 'll be judged best?
47. What they're saying shouldn't be repeated.
48. Give 'im five minutes and if he doesn't come, leave 'im.
49. Just 's I thought; he's not coming.
50. Coulda been the postman ringing the doorbell.

For purposes of indicating pronunciation, a number of quite innovative spellings are used in the items listed above. These were of course never shown to subjects, who worked exclusively from recorded oral interpretations of the sentences. For display purposes these could have been prepared in a phonemic or morphophonemic transcription. The choice of the forms cited above was decided by convenience; the sentences were read from a tapescript written with the contractions included as a reminder of how they should be pronounced.

It will be noticed that the criteria of administration are: 1) The test is paced; there have been no complaints that the pacing is too fast to allow time to consider and write answers; even students who tried all items and missed all of them had time to try. The timing was: instructions 3:05, first fifty 7:00, second fifty 7:15, with fifteen and ten second interludes, making a total test time of 17:45. 2) Subjects write items on a form that provides four columns of numbered blanks. Most words to be written are quite short; the actual lengths, in letters are:

1	2	3	4	5	6	7
2	18	14	9	6		1

3) Mistakes in spelling are not counted as errors if they are reasonable mistakes. Even errors like there for their were allowed. 4) There are no repeats; subjects must make their interpretations from a single rendition. 5) There is no rechecking of items possible, since no rerun of the item series is provided. The last two conditions (single presentation and no rerun) undoubtedly makes the test difficult, as shown by the fact no native speaker (or non-native) has gotten 100% on the test. Constant alertness for a sequence of 100 items is rather difficult. But the same difficult conditions are imposed on all

subjects, with a resulting rather considerable spread of scores, reflected in large standard deviations. The spread of scores for subjects has been:

Native	Non-native
83-99 (most over 90)	0-94 (rarely above 86)

For a typical administration subjects first listened to the introduction, after which the tape was stopped and questions were permitted. After these were answered, subjects were advised that there should be one word per blank (for each sentence), never more than one. Then the test sentences were played without interruption.

The test has been given several times. On November 18, 1974, a group of twenty Cubans, Puerto Ricans, and Chicanos working at the Dade County Spanish Curricula Development Center in Miami Beach were tested.[6] Their scores ranged from 5 to 94 with a mean of 49.8 and a predictably large standard deviation of 44.33. No correlations with other measures were possible, but native-English speaking colleagues confirmed from their own observations that the sequence of names developed by the test was closely representative of relative overall proficiency. Also the results seem to indicate a high degree of face validity when allowing for such factors as extent of contact with English (e.g. long residence on the U.S.-Mexico border), years of study and contact with English, and self-estimates of degree of bilingualism.

On November 20, 1974, the test was administered to sixteen third-year students of English at the Universidad Católica Madre y Maestra at Santiago de los Caballeros, Dominican Republic.[7] Scores ranged from 0 to 70, with a mean of 25.5 and a rather large standard deviation of 20.14. Fifteen of these scores (those for which comparison data are available) correlated very closely (r=.929) with course grades given for the same term, by four different teachers. This is an extremely high correlation, whose significance is perhaps related to the fact that there was a wide spread of test performance scores which proved to have predictive value with reference to course grades. Perhaps it is significant that all of the students were taught by and had close access to native English-speaking teachers, and the obscured pronunciations were therefore not unfamiliar.

On January 8, 1975, the test was administered to 44 fourth-year students of English at the University of Aleppo in Syria.[8] Scores ranged from 1 to 40 with a mean of 17.82 and a standard deviation of 12.64. Of these 44 subjects, two professors in the department of English were able to rate 36 on a seven-point scale, reflecting the criteria normally used at the university for grading. The correlation of test scores to these estimates was positive but not very strong: r = .4433. This is a substantially weaker correlation than the almost .93 noted for the Universidad Católica Madre y Maestra in the Dominican Republic. Two possible explanations are suggested: the smaller spread of scores (a range up to 70 at UCMM but only 40 at the University of Aleppo) indicating less competence overall and the fact that there are no native speakers of English on the staff at Aleppo (indeed none in the city as far as I could determine) and the students have far less opportunity for acquaintance with informal English such as is characterized by the obscured words and syllables in the experimental test. Perhaps even more important,

the test was given in Santiago in a language laboratory, where the acoustics were good, compared to a rather large amphitheater-type classroom in Aleppo.[9]

This raises the interesting question of what the test actually tests. Is it in fact a working knowledge of English grammar? Or is it listening comprehension? Or is it a measure of linguistic versatility or flexibility of verbal response--a kind of aptitude test or measure of linguistic ability? Or does it measure some level of general competence in English? Could it be considered a valid indirect measure of pronunciation?

An opportunity to provide tentative answers to at least some of these questions was made available by including the experimental test with the battery of tests given on December 17, 1974, to 81 applicants for admission to the American University in Cairo. The full battery consisted of Form H of the Michigan Test of English Language Proficiency (comprising three sections measuring vocabulary, grammar, and reading), a further test of listening comprehension, and a test of written composition.[10] Of this group of 81, 20 scored 80 or over, thus qualifying for admission to the university with no specific remedial English requirement, 22 scored 60 to 79 and were offered enrollment in the English Language Institute where they can hopefully increase their scores by means of special training in English language courses to 80 or better when retested later and thereby win admission as regular students, while 39 scored below 60 and were denied even provisional admission to the university. Students tested are mostly Egyptian, with a sprinkling of students from other Arabic-speaking countries, Europe, Africa, and Asia. Scores for the entire group on the integrative grammar test ranged from 0 to 84, with a mean of 24.9 and a standard deviation of 21.45.

With computer assistance[11] it was possible to make an analysis of the experimental test, the various scores and subscores of the admissions test, including the Michigan test, and the Thanawiyya Amma (the official Egyptian school leaving examination).[12] This analysis provides a considerable amount of information. Native speakers of Arabic formed the bulk of the students tested, 69 of the 81. In the experimental grammar test the Arabic speakers showed a ten percent improvement in the second half of the test, so some learning apparently takes place within the examination. The following data are illustrative:

	mean	standard deviation
First fifty items	10.2	9.81
Second fifty items	11.4	11.41
Full hundred items	21.6	20.99

Correlations are therefore cited for the second half of the test, with the first half considered practice. Actually the differences are very small, since the two halves correlate at .955.

The matrix in table 1 presents correlations for the complete data.

Table 1

	GR	VOC	RG	MICH	LC	WC	PLACE	TA	G1	G2	G3
GR	1.000	-	.863	.721	.927	.781	.901	.298	.767	.791	.789
VOC		1.000	-	-	-	-	-	-	-	-	-
RG			1.000	.714	.926	.734	.892	.331	.716	.746	.740
MICH				1.000	.824	.654	.777	.353	.614	.610	.619
LC					1.000	.826	.958	.407	.736	.760	.758
WC						1.000	.923	.413	.765	.777	.781
PLACE							1.000	.422	.795	.818	.817
TA								1.000	.234	.289	.266
G1									1.000	.955	.987
G2										1.000	.990
G3											1.000

The headings are GR grammar, VOC vocabulary, RG reading, MICH Michigan Test of English language proficiency (the combined scores of GR, VOC, AND RG), LC listening comprehension, WC written composition, PLACE placement (combined scores of MICH, LC, and WC), TA Thanawiyya Amma, G1 integrative grammar test, first 50 items, G2 second 50 items, G3 the combined 100 items. Unfortunately no figures were produced for VOC, though the VOC scores are included in the MICH figures.

A number of comments can be made on these correlations. The G2 scores correlate in descending order with PLACE .818, GR .791, WC .777, LC .760, RG .746, and MICH .610. The IGT (integrative grammar test) therefore appears to be a pretty good test of general ability and as an indirect measure of grammar, showing that it is indeed an integrative measure. This assumption is strengthened by the somewhat surprisingly high correlation with written composition, which in terms of the specific tasks imposed might seem to have little in common with the oral interpretation required by the IGT. But apparently the WC also measures general language competence.

The TA correlates only modestly with any of the other tests, which seems to confirm its status as a test of general achievement rather than as a language test. It would be interesting to compare the scores of the English subtest of the TA, if they were available, with the various sections of the proficiency and experimental tests. As it now functions the TA is a useful predictor of the probable scholastic success of students enrolling (better than the English tests), since it is a measure of general academic achievement.

A few comments on intercorrelations of the AUC proficiency test battery are pertinent. The PLACE scores (important because they represent the enrollment decision) correlate best with LC scores at .958, next best with WC at .923. These very high correlations should be discounted somewhat, since one third of the PLACE score comes respectively from LC and WC.[13] These two, however, correlate with each other at .826, a rate that exceeds correlation of either with G2 (at .760 and .777). The MICH score, the other one third of PLACE, correlates at .824 for LC, but only .654 for WC and an intermediate .776 for PLACE.

The GR subtest of the MICH test correlates strongly with LC at .927, well higher than the .791 with G2. GR and RG also correlate strongly at .863. Unfortunately, because of programming difficulties, correlations with VOC (alone, other

than when included in MICH) are unavailable; the experience of Irvine, Atai, and Oller suggests these would be lower--which is consistent with the assumption that integrative scores will correlate at higher levels than will discrete-point scores.

What do these correlations mean? They seem to indicate that a useful product can be obtained from an integrative grammar test of the type described in this article, that it does yield a score that compares well with more traditional forms of testing. The IGT has the advantage of being efficient (about 18 minutes to administer compared to the two and a half hours for the proficiency battery) and can be scored easily and rapidly (though machine scoring in its present form is not feasible). This of course does not automatically suggest the replacement of the present testing procedure by the IGT, which is still very experimental. This is especially true because changes in testing policy and instruments often produces unpredictable results. A battery of varied form encourages broad linguistic preparation that may disappear if the test is narrowed. In Ethiopia, for instance, many teachers discontinued the teaching of writing skills when the written composition was removed from the secondary school-leaving examination, with predictable and unfortunate consequences. I would be very alarmed at a program of test review which deemphasized grammar, reading, and writing in order to concentrate on the skill of unravelling the distortions normal to obscured forms in informal speech in English.

Perhaps the most that can be said is that further research seems justified. Maybe a proficiency battery could profitably include a subtest on the analysis and interpretation of obscured forms. Or maybe further experimentation will send this activity back to the classroom as a drill in listening comprehension for students who will need to communicate orally and informally with native speakers of English.

REFERENCES

Carroll, John B. "Fundamental Considerations in Testing for English Language Proficiency of Foreign Students," Testing, Washington: Center for Applied Linguistics, 1961, pp. 31-40. Reprinted in Teaching English as a Second Language: A Book of Readings, 2nd edition, Harold B. Allen and Russell N. Campbell, eds., New York: McGraw-Hill, 1972, pp. 313-321.

Cohen, Andrew D. "An Up-date Report on the English Language Placement Exam," UCLA Workpapers in English as a Second Language, 7 (1973) pp. 101-105.

Harris, David P. Testing English as a Second Language. New York: McGraw-Hill, 1969.

Irvine, Patricia, Parvin Atai, and John W. Oller, Jr. "Cloze, Dictation, and the Test of English as a Foreign Language," Language Learning, 24:2 (December, 1974) pp. 245-252.

Lee, Richard R. "A Taxonomy of Second Language Behaviors" TESOL Quarterly 6:3 (September, 1972) pp. 209-220.

Oller, John W., Jr. "Discrete-Point Tests Versus Tests of Integrative Skills," Focus on the Learner: Pragmatic Perspectives for the Language Teacher, ed. by John W. Oller, Jr., and Jack C. Richards. Rowley: Newbury House, 1973a, pp. 184-199.

Oller, John W., Jr. "Integrative and Discrete Point Tests at UCLA: A Reply," UCLA Workpapers in English as a Second Language 7 (1973b) pp. 85-94.

Rand, Earl J. "Integrative and Discrete-Point Tests at UCLA," UCLA Workpapers in English as a Second Language 6 (1972) pp. 67-78.

Rand, Earl J. "Comments on a Reply," UCLA Workpapers in English as a Second Language 7 (1973) pp. 95-99.

NOTES

[1]Rand 1972, 1973 and Oller 1973a and 1973b. See also Carroll 1961.

[2]Oller 1973b, p. 198.

[3]Cf. Cohen, p. 104.

[4]Lee 1972, p. 213.

[5]My thanks to Father Albert Linders, S.J., who arranged for the administration of the test at the Collège de la Sainte Famille, a Catholic school for boys in Cairo.

[6]My thanks to Mrs. Denise Mesa, staff writer for the Spanish Curricula Development Center, for arranging this administration.

[7]My thanks to Professor Gari Browning of the Department of English, Universidad Católica Madre y Maestra for arranging this administration.

[8]My thanks to Professors Hashem A. Madi and Louis Morcos of the Department of English, University of Aleppo for facilitating this administration of the test.

[9]When a research result is both unexpected and disappointing, there is a strong temptation to try and explain. Why were the correlations in Santiago so high and in Aleppo so low? Were there perhaps better judgments in the Santiago course grades than in the Aleppo teacher ratings (though course grades are surely ratings by teachers)? Course grades may be more seriously prepared, considering evidence from examinations, written assignments, etc. that may not be prominent in an off-the-cuff evaluation. As an example of less systematic consideration it is noted that in Aleppo one teacher was sufficiently well acquainted with only 28 and the other with 33 of the 44 students to be willing to offer a rating. Also in almost a dozen cases where both teachers evaluated the same student the ratings were disparate, suggesting the possibility that some names and students in the large class may have been confused. Possibly if course grades could be obtained, the correlation would rise to the 50's, which Harris indicates is an acceptable level for such subjective external criteria as supervisor ratings (Harris, 1969, p. 20).

[10]The Michigan test consists of three sections with a total of 100 items (multiple-choice items - four choices) distributed as follows: grammar 40, vocabulary 40, reading comprehension 20, with a time allowance of 75 minutes provided. Additionally a test of listening comprehension of 90 items (multiple choice - three choices) is administered in the language laboratory, taking just under 25 minutes. Finally there is a writing composition test, the students writing one selection on an assigned subject, for which 30 minutes is allowed. The equated scores of the Michigan test of grammar, vocabulary, and reading comprehension), the listening comprehension test, and the composition test, with scores expressed as precentages, each count one-third toward the overall score for the full examination. All tests are hand scored by key except the compositions, which are read blind (with examinee's name removed), by at least three members of the teaching staff, and as many as five if the first three readers do not agree.

[11]I would like to acknowledge, with appreciation, the assistance of the Computer Center, American University in Cairo and the programming assistance of Mr. Mokbel H. Fayyad.

[12]The Thanawiyya Amma is a series of general achievement tests, the culmination of the third (and last) year of secondary school as a school-leaving examination. Another function is qualification for enrollment in the national universities and higher institutes. All subjects taken in the third year of secondary school are examined and a composite score is calculated. The English subtest takes various forms depending on whether it is prepared for the scientific or literary stream, whether it is tested as a first or second foreign language, etc. A reasonably typical format is two papers of two sections each for a total of forty points distributed over written composition for ten points (expository writing on an assigned topic), reading comprehension for ten points (sentence replies to questions on cited passages), set texts for ten points (literature and where appropriate science texts to be interpreted), and a final section divided between language or grammar for five points (requiring various kinds of simple sentence transformations) and translation for five points (rendering an English text into Arabic and an Arabic text into English). A passing score is thirty per cent, qualifying a candidate for secondary school graduation, though a much higher score, variously set for different institutions, is required for admission to higher courses.

[13]It would be desirable to follow the suggestion of Irvine, Atai, and Oller (1974) to calculate correlations omitting in turn each subtest from the total scores before making comparisons, to avoid the otherwise inevitable inflation of correlation coefficients when a score is compared in part to itself.